Uncertainty Relations and Sparse Signal Recovery for Pairs of General Signal Sets

Patrick Kuppinger, *Student Member, IEEE*, Giuseppe Durisi, *Member, IEEE*, and Helmut Bölcskei, *Fellow, IEEE*

Abstract

We present an uncertainty relation for the representation of signals in two different general (possibly redundant or incomplete) signal sets. This uncertainty relation is relevant for the analysis of signals containing two distinct features each of which can be described sparsely in a suitable general signal set. Furthermore, the new uncertainty relation is shown to lead to improved sparsity thresholds for recovery of signals that are sparse in general dictionaries. Specifically, our results improve on the well-known $(1+1/d)/2$-threshold for dictionaries with coherence d by up to a factor of two. Furthermore, we provide probabilistic recovery guarantees for pairs of general dictionaries that also allow us to understand which parts of a general dictionary one needs to randomize over to "weed out" the sparsity patterns that prohibit breaking the square-root bottleneck.

I. Introduction and Outline

A milestone in the sparse signal recovery literature is the uncertainty relation for the Fourier-identity pair found in [1]. This uncertainty relation was extended to pairs of arbitrary orthonormal bases (ONBs) in [2]. Besides being interesting in their own right, these uncertainty relations are fundamental in the formulation of recovery guarantees for signals that contain two distinct features, each of which can be described sparsely using an ONB. If the individual features are, however, sparse only in overcomplete signal sets (i.e., in frames [3]), the two-ONB result [1], [2] cannot be applied. The goal of this paper is to find uncertainty relations and corresponding signal recovery guarantees for signals that are sparse in pairs of general (possibly redundant) signal sets. Redundancy in the individual signal sets allows us to succinctly describe a wider class of features. Concrete examples for this setup can be found in the feature extraction or morphological component analysis literature (see, e.g., [4], [5] and references therein).

P. Kuppinger and H. Bölcskei are with the Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland, Email: {patricku,boelcskei}@nari.ee.ethz.ch

G. Durisi is with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, Email: durisi@chalmers.se

The material in this paper was presented in part at the IEEE Information Theory Workshop (ITW), Taormina, Italy, October 2009, and at the IEEE International Symposium on Information Theory (ISIT), Austin, TX, June 2010.

In order to put our results into perspective and to detail our contributions, we first briefly recapitulate the formal setup considered in the sparse signal recovery literature [6], [7], [2], [8]–[11].

A. *Sparse Signal Recovery Methods*

Consider the problem of recovering unknown vectors from small numbers of linear non-adaptive measurements. More formally, let $\mathbf{x} \in \mathbb{C}^N$ be an unknown vector that is observed through a measurement matrix $\mathbf{D}$ with columns[1] $\mathbf{d}_i \in \mathbb{C}^M$, $i = 1, 2, \ldots, N$, according to

$$\mathbf{y} = \mathbf{D}\mathbf{x}$$

where $\mathbf{y} \in \mathbb{C}^M$ and $M \ll N$. If we do not impose additional assumptions on $\mathbf{x}$, the problem of recovering $\mathbf{x}$ from $\mathbf{y}$ is obviously ill-posed. The situation changes drastically if we assume that $\mathbf{x}$ is sparse in the sense of having only a few nonzero entries. More specifically, let $\|\mathbf{x}\|_0$ denote the number of nonzero entries of $\mathbf{x}$, then

$$\text{(P0)} \quad \text{minimize } \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{D}\mathbf{x}$$

can recover $\mathbf{x}$ without prior knowledge of the positions of the nonzero entries of $\mathbf{x}$. Equivalently, we can interpret (P0) as the problem of finding the sparsest representation of the vector $\mathbf{y}$ in terms of the "dictionary elements" (columns) $\mathbf{d}_i$. In this context, the matrix $\mathbf{D}$ is often referred to as dictionary.

Since (P0) is an NP-hard problem [12] (it requires a combinatorial search), it is computationally infeasible, even for moderate problem sizes N, M. Two popular and computationally more tractable alternatives to solving (P0) are *basis pursuit* (BP) [13], [6]–[8], [2], [9] and *orthogonal matching pursuit* (OMP) [14], [15], [9]. BP is a convex relaxation of the (P0) problem, namely

$$\text{(BP)} \quad \text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{y} = \mathbf{D}\mathbf{x}.$$

Here, $\|\mathbf{x}\|_1 = \sum_i |x_i|$ denotes the ℓ_1-norm of the vector $\mathbf{x}$. OMP is an iterative greedy algorithm that constructs a sparse representation of $\mathbf{y}$ by selecting, in each iteration, the column of $\mathbf{D}$ most "*correlated*" with the difference between $\mathbf{y}$ and its current approximation.

Two questions that arise naturally are: 1) Under which conditions is $\mathbf{x}$ the unique solution of (P0)? 2) Under which conditions is this solution delivered by BP and/or OMP? Answers to these questions are typically expressed in terms of sparsity thresholds on the unknown vector $\mathbf{x}$ [6]–[8], [2], [9]. These sparsity thresholds either hold for all possible sparsity patterns and values of nonzero entries in $\mathbf{x}$, in which case we speak of *deterministic* sparsity thresholds. Alternatively, one may be interested in so-called *probabilistic* or—following the terminology used in [10]—*robust* sparsity thresholds, which

[1] Throughout the paper, we shall assume that the columns of $\mathbf{D}$ span $\mathbb{C}^M$ and have unit ℓ_2-norm.

hold for *most* sparsity patterns and values of nonzero entries in $\mathbf{x}$. Intuitively, robust sparsity thresholds are larger than deterministic ones. More precisely, as the number of measurements M grows large, deterministic sparsity thresholds generally scale at best as $\sqrt{M}$. Robust sparsity thresholds, in contrast, break this *square-root bottleneck*. In particular, they scale on the order of $M/(\log N)$ [11]. However, this comes at a price: Uniqueness of the solution of[2] (P0) and recoverability of the (P0)-solution through BP is guaranteed only with high probability with respect to the choice of[3] $\mathbf{x}$.

Both deterministic and probabilistic sparsity thresholds are typically expressed in terms of the dictionary *coherence*, defined as the maximum absolute value over all inner products between pairs of distinct columns of $\mathbf{D}$.

An alternative approach is to assume that the dictionary $\mathbf{D}$ is random (rather than the vector $\mathbf{x}$) and to determine thresholds that hold for all (sufficiently) sparse $\mathbf{x}$ with high probability with respect to the choice of $\mathbf{D}$ [17]–[19]. Throughout this paper, we consider deterministic dictionaries exclusively.

Note that when considering signals that consist of two distinct features, each of which can be described sparsely using an ONB [2], [6], [20], [9], the corresponding dictionary $\mathbf{D}$ is given by the concatenation of these two ONBs. One obvious way of obtaining recovery guarantees for signals that are sparse in pairs of general signal sets is to *concatenate* these general signal sets, view the concatenation as one (general) dictionary, and apply the sparsity thresholds for general dictionaries reported in, e.g., [7]–[9], [11]. However, these sparsity thresholds depend only on the coherence of the resulting overall dictionary $\mathbf{D}$ and, in particular, do not take into account the coherence parameters of the two constituent signal sets.

In this paper, we show that the sparsity thresholds can be significantly improved not only if $\mathbf{D}$ is the concatenation of two ONBs—as was done in [2], [8], [20], [9]—but also if $\mathbf{D}$ consists of the concatenation of two general signal sets (or *sub-dictionaries*) with known coherence parameters.

B. Contributions

Our contributions can be detailed as follows. Based on a novel uncertainty relation for pairs of general (redundant or incomplete) signal sets, we obtain a novel deterministic sparsity threshold guaranteeing (P0)-uniqueness for dictionaries that are given by the concatenation of two general sub-dictionaries with known coherence parameters. Additionally, we derive a novel threshold guaranteeing that BP and OMP recover this unique (P0)-solution. Our thresholds improve significantly on the known deterministic sparsity thresholds one would obtain if the concatenation of two sub-dictionaries were viewed as a general dictionary, thereby ignoring the additional information about the sub-dictionaries' coherence

[2]Whenever we speak of uniqueness of the solution of (P0), we mean that the unique solution of (P0) applied to $\mathbf{y} = \mathbf{Dx}$ is given by $\mathbf{x}$.

[3]Robust sparsity thresholds for OMP to deliver the unique (P0)-solution are still unknown. For the *multichannel* scenario, first results along these lines were reported in [16], where it is shown that the probability of reconstruction error decays exponentially with the number of channels.

parameters. More precisely, this improvement can be up to a factor of two. Moreover, the known sparsity thresholds for general dictionaries and the ones for the concatenation of two ONBs follow from our results for the concatenation of general sub-dictionaries as special cases.

Concerning probabilistic sparsity thresholds for the concatenation of two general dictionaries, we address the following question: Given a *general* dictionary, can we break the square-root bottleneck while only randomizing the sparsity patterns over a certain part of the overall dictionary? By extending the known results for the two-ONB setting [10], [11] to the concatenation of two general dictionaries, we show that the answer is in the affirmative. Our results allow us to identify parts of a general dictionary the sparsity patterns need to be randomized over so as to break the square-root bottleneck.

C. Notation

We use lowercase boldface letters for column vectors, e.g., $\mathbf{x}$, and uppercase boldface letters for matrices, e.g., $\mathbf{D}$. For a given matrix $\mathbf{D}$, we denote its ith column by $\mathbf{d}_i$, its conjugate transpose by $\mathbf{D}^H$, and its Moore-Penrose inverse by $\mathbf{D}^\dagger$. Slightly abusing notation, we say that $\mathbf{d} \in \mathbf{D}$ if $\mathbf{d}$ is a column of the matrix $\mathbf{D}$. The spectral norm of a matrix $\mathbf{D}$ is $\|\mathbf{D}\| = \sqrt{\lambda_{\max}(\mathbf{D}^H\mathbf{D})}$, where $\lambda_{\max}(\mathbf{D}^H\mathbf{D})$ denotes the maximum eigenvalue of $\mathbf{D}^H\mathbf{D}$. The minimum and maximum singular value of $\mathbf{D}$ are denoted by $\sigma_{\min}(\mathbf{D})$ and $\sigma_{\max}(\mathbf{D})$, respectively; $\operatorname{rank}(\mathbf{D})$ stands for the rank of $\mathbf{D}$, $\|\mathbf{D}\|_{1,2} = \max_i\{\|\mathbf{d}_i\|_2\}$, and $\|\mathbf{D}\|_{1,1} = \max_i\{\|\mathbf{d}_i\|_1\}$. The smallest eigenvalue of the positive-semidefinite matrix $\mathbf{G}$ is denoted by $\lambda_{\min}(\mathbf{G})$. We use $\mathbf{I}_n$ to refer to the $n \times n$ identity matrix; $\mathbf{0}_{m,n}$ and $\mathbf{1}_{m,n}$ stand for the all-zero and all-one matrix of size $m \times n$, respectively. We denote the n-dimensional all-ones and all-zeros column vector by $\mathbf{1}_n$ and $\mathbf{0}_n$, respectively. The natural logarithm is referred to as $\log$. The set of all positive integers is $\mathbb{N}^+$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = \Omega(g(x))$ means that there exists a real number x_0 such that $|f(x)| \geq k_1|g(x)|$ for all $x > x_0$, where k_1 is a finite constant. The notation $f(x) = \mathcal{O}(g(x))$ means that there exists a real number x_0 such that $|f(x)| \leq k_2|g(x)|$ for all $x > x_0$, where k_2 is a finite constant. Furthermore, we write $f(x) = \Theta(g(x))$ if there exists a real number x_0 and finite constants k_1 and k_2 such that $k_1|g(x)| \leq |f(x)| \leq k_2|g(x)|$ for all $x > x_0$. For $u \in \mathbb{R}$, we define $[u]^+ = \max\{0, u\}$. Whenever we say that a vector $\mathbf{x} \in \mathbb{C}^N$ has a *randomly* chosen sparsity pattern of cardinality L, we mean that the support set of $\mathbf{x}$ (i.e., the set of nonzero entries of $\mathbf{x}$) is chosen uniformly at random among all $\binom{N}{L}$ possible support sets of cardinality L.

II. Deterministic Sparsity Thresholds

A. A Brief Review of Relevant Previous Work

A quantity that is intimately related to the uniqueness of the solution of (P0) is the *spark* of a dictionary $\mathbf{D}$, defined as the smallest number of linearly dependent columns of $\mathbf{D}$ [7]. More specifically, the

following result holds [7], [8]: For a given dictionary $\mathbf{D}$ and measurement outcome $\mathbf{y} = \mathbf{D}\mathbf{x}$, the unique solution of (P0) is given by $\mathbf{x}$ if

$$\|\mathbf{x}\|_0 < \frac{\mathrm{spark}(\mathbf{D})}{2}. \tag{1}$$

Unfortunately, determining the spark of a dictionary is an NP-hard problem, i.e., a problem that is as hard as solving (P0) directly. It is possible, though, to derive easy-to-compute lower bounds on $\mathrm{spark}(\mathbf{D})$ that are explicit in the coherence of $\mathbf{D}$ defined as

$$d = \max_{i \neq j} \left| \mathbf{d}_i^H \mathbf{d}_j \right|. \tag{2}$$

We next briefly review these lower bounds. Let us first consider the case where $\mathbf{D}$ is the concatenation of two ONBs. Denote the set of all dictionaries that are the concatenation of two ONBs and have coherence d by $\mathcal{D}_{\text{onb}}(d)$. It was shown in [2] that for $\mathbf{D} \in \mathcal{D}_{\text{onb}}(d)$, we have

$$\mathrm{spark}(\mathbf{D}) \geq \frac{2}{d}. \tag{3}$$

Substituting (3) into (1) yields the following sparsity threshold guaranteeing that the unique solution of (P0) applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$:

$$\|\mathbf{x}\|_0 < \frac{1}{d}. \tag{4}$$

Furthermore, it was shown in [2], [20], [9] that for this unique solution to be recovered by BP and OMP it is sufficient to have

$$\|\mathbf{x}\|_0 < \frac{\sqrt{2} - 0.5}{d} \approx \frac{0.9}{d}. \tag{5}$$

A question that arises naturally is: What happens if the dictionary $\mathbf{D}$ is not the concatenation of two ONBs? There exist sparsity thresholds in terms of d for general dictionaries. Specifically, let us denote the set of all dictionaries with coherence d by $\mathcal{D}_{\text{gen}}(d)$. It was shown in [7]–[9] that for $\mathbf{D} \in \mathcal{D}_{\text{gen}}(d)$ we have

$$\mathrm{spark}(\mathbf{D}) \geq 1 + \frac{1}{d}. \tag{6}$$

Using (6) in (1) yields the following sparsity threshold guaranteeing that the unique solution of (P0) applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$:

$$\|\mathbf{x}\|_0 < \frac{1}{2}\left(1 + \frac{1}{d}\right). \tag{7}$$

Interestingly, one can show that (7) also guarantees that BP and OMP recover the unique (P0)-solution [7]–[9].

The set $\mathcal{D}_{\text{gen}}(d)$ is large, in general, and contains a variety of structurally very different dictionaries, ranging from equiangular tight frames (where the absolute values of the inner products between any two distinct dictionary elements are equal) to dictionaries where the maximum inner product is achieved by

one pair only. The sparsity threshold in (7) is therefore inevitably rather crude. Better sparsity thresholds are possible if one considers subsets of $\mathcal{D}_{\text{gen}}(d)$, such as, e.g., $\mathcal{D}_{\text{onb}}(d) \subset \mathcal{D}_{\text{gen}}(d)$. A dictionary $\mathbf{D} \in \mathcal{D}_{\text{onb}}(d)$ also satisfies $\mathbf{D} \in \mathcal{D}_{\text{gen}}(d)$, and, hence, the sparsity threshold in (7) applies. However, the additional structural information about $\mathbf{D}$ being the concatenation of two ONBs, i.e., $\mathbf{D} \in \mathcal{D}_{\text{onb}}(d)$, allows us to obtain the improved sparsity thresholds in (4) and (5), which are (for $d \ll 1$) almost a factor of two higher (better) than the threshold in (7). As a side remark, we note that the threshold for the two-ONB case in (5) drops below that in (7), valid for general dictionaries, if $d > 2(\sqrt{2}-1)$. This is surprising as exploiting structural information should lead to a higher sparsity threshold. We will show, in Section II-B, that one can refine the threshold in (5) so as to fix this problem.

B. Novel Deterministic Sparsity Thresholds for the Concatenation of Two General Signal Sets

We consider dictionaries with coherence d that consist of two sub-dictionaries with coherence a and b, respectively. The set of all such dictionaries will be denoted as $\mathcal{D}(d,a,b)$. A dictionary $\mathbf{D} \in \mathcal{D}(d,a,b)$ of dimension $M \times N$ (with $N \geq M$) can be written as $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$, where the sub-dictionary $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ has coherence a and the sub-dictionary $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ has coherence b. We remark that the two sub-dictionaries need not be ONBs, need not have the same number of elements and need not span $\mathbb{C}^M$, but their concatenation is assumed to span $\mathbb{C}^M$. Without loss of generality, we assume, throughout the paper, that $a \leq b$. For fixed d,[4] we have that $\mathcal{D}(d,a,b) \subset \mathcal{D}_{\text{gen}}(d)$. Hence, we consider subsets $\mathcal{D}(d,a,b)$ of the set $\mathcal{D}_{\text{gen}}(d)$ parametrized by the coherence parameters a and b.

For $\mathbf{D} \in \mathcal{D}(d,a,b)$ we derive sparsity thresholds in terms of d, a, and b and show that these thresholds improve upon those in (7) for general dictionaries $\mathbf{D} \in \mathcal{D}_{\text{gen}}(d)$. This improvement is a result of the restriction to a subset of dictionaries in $\mathcal{D}_{\text{gen}}(d)$, namely $\mathcal{D}(d,a,b)$, and of exploiting the additional structural information (in terms of the coherence parameters a and b) available about dictionaries $\mathbf{D}$ in this subset.

Every dictionary in $\mathcal{D}_{\text{gen}}(d)$ can be viewed as the concatenation of two sub-dictionaries. Our results therefore state that viewing a dictionary $\mathbf{D} \in \mathcal{D}_{\text{gen}}(d)$ as the concatenation of two sub-dictionaries leads to improved sparsity thresholds provided that the coherence parameters a and b of the respective sub-dictionaries are available. Moreover, the improvements will be seen to be up to a factor of two if a and b are sufficiently small.

The sparsity threshold for uniqueness of the solution of (P0) for dictionaries $\mathbf{D} \in \mathcal{D}(d,a,b)$, formalized in Theorem 2 below, is based on a novel uncertainty relation for pairs of general dictionaries, stated in the following lemma.

[4]We assume throughout the paper that $d > 0$. For $d = 0$ the dictionary $\mathbf{D}$ consists of orthonormal columns, and, hence, every unknown vector $\mathbf{x}$ can be uniquely recovered from the measurement outcome $\mathbf{y}$ according to $\mathbf{x} = \mathbf{D}^H\mathbf{y}$.

Lemma 1: Let $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ be a dictionary with coherence a, $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ a dictionary with coherence b, and denote the coherence of the concatenated dictionary $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$, $\mathbf{D} \in \mathbb{C}^{M \times N}$, by d. For every vector $\mathbf{s} \in \mathbb{C}^M$ that can be represented as a linear combination of n_a columns of $\mathbf{A}$ and, equivalently, as a linear combination of n_b columns of $\mathbf{B}$,[5] the following inequality holds:

$$n_a n_b \geq \frac{[1 - a(n_a - 1)]^+ \, [1 - b(n_b - 1)]^+}{d^2}. \tag{8}$$

Proof: See Appendix A. ■

The uncertainty relation for the union of two-ONB case derived in [2] is a special case of (8). In particular, if $a = b = 0$, then (8) reduces to the result reported in [2, Thm. 1]:

$$n_a n_b \geq \frac{1}{d^2}. \tag{9}$$

Note that, differently from [2, Thm. 1], the lower bound in (9) holds not only for the concatenation of two ONBs, but also for the concatenation of two sub-dictionaries $\mathbf{A}$ and $\mathbf{B}$ that contain orthonormal columns but individually do not necessarily span $\mathbb{C}^M$ (but their concatenation spans $\mathbb{C}^M$). Lemma 1 allows us to easily recover several other well-known results such as, e.g., the well-known lower bound in (6) on the spark of a dictionary. To see this note that when $n_b = 0$ in Lemma 1 (and thus $\mathbf{s} = \mathbf{0}_M$, by definition) then the n_a columns in $\mathbf{A}$ participating in the representation of $\mathbf{s}$ are linearly dependent. Moreover, for $n_b = 0$ we have $[1 - b(n_b - 1)]^+ = (1 + b) > 0$. Therefore, it follows from (8) that necessarily $[1 - a(n_a - 1)]^+ = 0$ and thus $n_a \geq 1 + 1/a$, which agrees with the lower bound on the spark of the (sub-)dictionary $\mathbf{A}$ [7]–[9]. A similar observation follows for $n_a = 0$.

More importantly, Lemma 1 also allows us to derive a new lower bound on the spark of the overall dictionary $\mathbf{D} = [\mathbf{A}\ \mathbf{B}] \in \mathcal{D}(d, a, b)$. When used in (1), this result then yields a new sparsity threshold guaranteeing uniqueness of the (P0)-solution. We show that this threshold improves upon that in (7), which would be obtained if we viewed $\mathbf{D}$ simply as a general dictionary in $\mathcal{D}_{\text{gen}}(d)$, thereby ignoring the fact that the dictionary under consideration belongs to a subset, namely $\mathcal{D}(d, a, b)$, of $\mathcal{D}_{\text{gen}}(d)$.

Theorem 2: For $\mathbf{D} \in \mathcal{D}(d, a, b)$, a sufficient condition for the vector $\mathbf{x}$ to be the unique solution of (P0) applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is that

$$\|\mathbf{x}\|_0 < \frac{f(\hat{x}) + \hat{x}}{2} \tag{10}$$

where

$$f(x) = \frac{(1 + a)(1 + b) - xb(1 + a)}{x(d^2 - ab) + a(1 + b)}$$

and $\hat{x} = \min\{x_b, x_s\}$. Furthermore,

$$x_b = \frac{1 + b}{b + d^2}$$

[5]For $n_a = 0$ or $n_b = 0$ we define $\mathbf{s} = \mathbf{0}_M$. We exclude the trivial case $n_a = n_b = 0$.

and

$$x_s = \begin{cases} \dfrac{1}{d}, & \text{if} \quad a = b = d, \\ \\ \dfrac{d\sqrt{(1+a)(1+b)} - a - ab}{d^2 - ab}, & \text{otherwise.} \end{cases}$$

Proof: See Appendix B. ■

The sparsity threshold in (10) reduces to that in (7) when $b = d$ (irrespective of a) or when $d = 1$ (irrespective of a and b). Hence, the sparsity threshold in (10) does not improve upon that in (7) if the pair of columns achieving the overall dictionary coherence d appears in the same sub-dictionary $\mathbf{B}$ (recall that we assumed $b \geq a$), or if $d = 1$. In all other cases, the sparsity threshold in (10) can be shown to be strictly larger than that in (7). This result is proven in Appendix C. The improvement can be up to a factor of two. We demonstrate this for the special case $a = b$, for which the sparsity threshold in (10) takes a particularly simple form. In this case, as can easily be verified, $x_s \leq x_b$ so that (10) reduces to

$$\|\mathbf{x}\|_0 < \frac{1+b}{d+b}. \tag{11}$$

For $a = b = 0$ the sparsity threshold in (11) reduces to the known sparsity threshold for dictionaries in $\mathcal{D}_{\text{onb}}(d)$ specified in (4). Note, however, that the threshold in (11) with $b = 0$ holds for all $\mathbf{D} \in \mathcal{D}(0, 0, d)$, thereby also including sub-dictionaries $\mathbf{A}$ and $\mathbf{B}$ that contain orthonormal columns but do not necessarily individually span $\mathbb{C}^M$ (but their concatenation spans $\mathbb{C}^M$). Setting $b = \epsilon d$ with $\epsilon \in [0, 1]$ and noting that for $d \ll 1$ the ratio between the sparsity threshold in (11) and that in (7) is roughly $2/(1+\epsilon)$, which for $\epsilon \ll 1$ is almost two. Note that, for small coherence parameters a and b, the elements in each of the two sub-dictionaries $\mathbf{A}$ and $\mathbf{B}$ are close to being orthogonal to each other. Fig. 1 shows the sparsity threshold in (11) for $d = 0.01$ as a function of b. We can see that for $b \ll d$ the threshold in (11) is, indeed, almost a factor of two larger than that in (7).

So far, we focused on thresholds guaranteeing (P0)-uniqueness. We next present thresholds guaranteeing recovery of the unique (P0)-solution via BP and OMP for dictionaries $\mathbf{D} \in \mathcal{D}(d, a, b)$. The recovery conditions we report in Theorem 3 and Corollary 4 below, depend on b and d, but not on a. Slightly improved thresholds that also depend on a can be derived following similar ideas as in the proofs of Theorem 3 and Corollary 4. The resulting expressions are, however, unwieldy and will therefore not be presented here.

Theorem 3: Suppose that $\mathbf{y} \in \mathbb{C}^M$ can be represented as $\mathbf{y} = \mathbf{D}\mathbf{x}$, where $\mathbf{x}$ has n_a nonzero entries corresponding to columns of $\mathbf{A}$ and n_b nonzero entries corresponding to columns of $\mathbf{B}$. Without loss of generality, we assume that $n_a \leq n_b$. A sufficient condition for BP *and* OMP to recover $\mathbf{x}$ is

$$2n_a(1+b)b + n_b(1+b)(d+b) + 2n_a n_b(d^2 - b^2) < (1+b)^2. \tag{12}$$

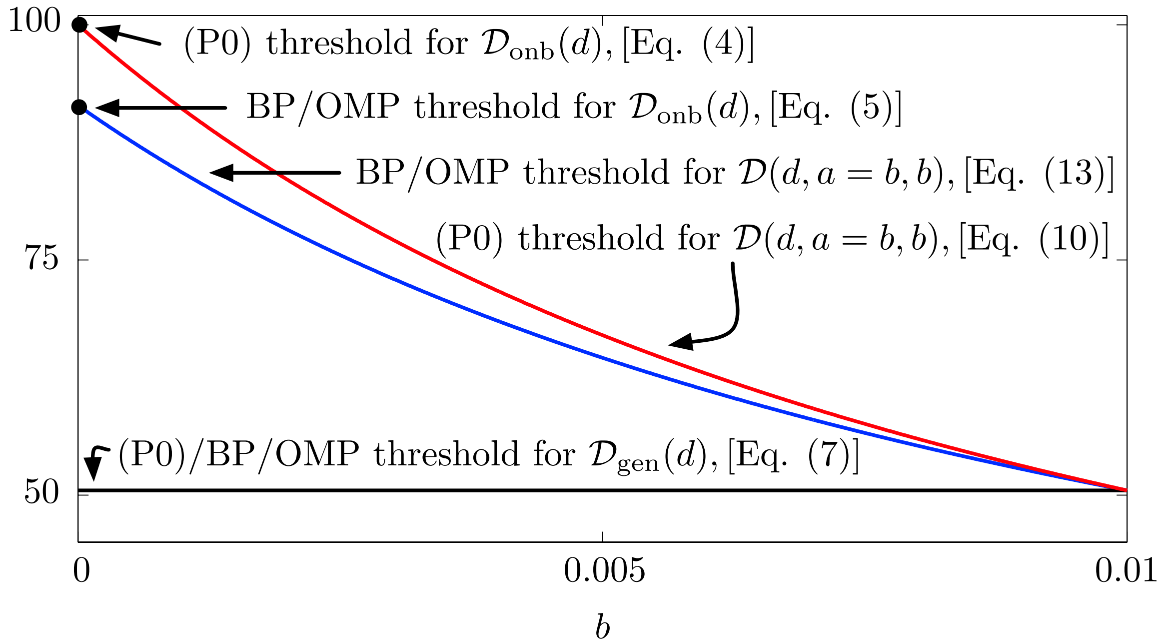

Figure 1. Deterministic sparsity thresholds guaranteeing uniqueness of (P0) and recoverability via BP and OMP for dictionaries in $\mathcal{D}_{\text{gen}}(d)$, $\mathcal{D}_{\text{onb}}(d)$, and $\mathcal{D}(d, a, b)$. We set $d = 0.01$ and consider the special case $a = b$. Note that for $a = b$, the threshold in (10) reduces to that in (11).

Proof: See Appendix D. ■

Theorem 3 generalizes the result in [2, Sec. 6], [9, Cor. 3.8] for the concatenation of two ONBs to dictionaries $\mathbf{D} \in \mathcal{D}(d, a, b)$. In particular, (12) reduces to [9, Eq. (16)] when $b = 0$ (since $a \leq b$, this implies $a = 0$). Furthermore, when $b = d$, the condition in (12) simplifies to $n_a + n_b < (1 + 1/d)/2$, thereby recovering the sparsity threshold in (7). Thus, if the pair of columns achieving the overall dictionary coherence is in the same sub-dictionary $\mathbf{B}$ (recall that we assumed $b \geq a$), no improvement over the well-known $(1 + 1/d)/2$-threshold for dictionaries in $\mathcal{D}_{\text{gen}}(d)$ is obtained. Theorem 3 depends explicitly on n_a and n_b. In the following corollary, we provide a recovery guarantee in the form of a sparsity threshold that depends on n_a and n_b only through the overall sparsity level of $\mathbf{x}$ according to $\|\mathbf{x}\|_0 = n_a + n_b$.

Corollary 4: For $\mathbf{D} \in \mathcal{D}(d, a, b)$ a sufficient condition for BP *and* OMP to deliver the unique solution of (P0) is

$$\|\mathbf{x}\|_0 < \begin{cases} \dfrac{(1+b)\big[\xi - (d+3b)\big]}{2(d^2 - b^2)}, & \text{if } b < d \text{ and } \kappa(d, b) > 1, \\[2ex] \dfrac{1 + 2d^2 + 3b - d(1+b)}{2(d^2 + b)}, & \text{otherwise} \end{cases} \tag{13}$$

with

$$\kappa(d,b) = \frac{(1+b)(\xi - 4b)}{4(d^2 - b^2)} \tag{14}$$

and $\xi = 2\sqrt{2}\sqrt{d(b+d)}$.

Proof: See Appendix E. ■

The sparsity threshold in (13) reduces to the sparsity threshold in (7) when $b = d$ or when $d = 1$ (irrespective of b). In all other cases, the sparsity threshold in (13) is strictly larger than that in (7) (see Appendix F). The threshold in (13) is complicated as we have to deal with two different cases. The distinction between these two cases is, however, crucial to ensure that the threshold in (13) does not fall below that in (7).[6] It turns out that the first case in (13) is active whenever $b < d < 3/5$, which covers essentially all practically relevant cases. In fact, for dictionaries with coherence $d \geq 3/5$, the sparsity threshold in (13) allows for at most one nonzero entry in $\mathbf{x}$.

The improvement of the sparsity threshold in (13) over that in (7) can be up to a factor of almost two. This can be seen by setting $b = \epsilon d$ with $\epsilon \in [0,1)$ and noting that for $d \ll 1$ the ratio between the sparsity threshold in the first case in (13) and that in (7) is roughly $(2\sqrt{2(1+\epsilon)} - (1+3\epsilon))/(1-\epsilon^2)$, which for $\epsilon \ll 1$ is approximately 1.8. Fig. 1 shows the threshold in (13) for $d = 0.01$ as a function of b. We can see that for $b \ll d$ the threshold in (13) is, indeed, almost a factor of two larger than that in (7).

If $\mathbf{D}$ is the concatenation of two ONBs, and hence $a = b = 0$, the sparsity threshold in (13) reduces to

$$\|\mathbf{x}\|_0 < \begin{cases} \dfrac{\sqrt{2} - 0.5}{d}, & \text{if } d < \dfrac{1}{\sqrt{2}}, \\[2ex] 1 + \dfrac{1-d}{2d^2}, & \text{otherwise.} \end{cases} \tag{15}$$

For $d < 1/\sqrt{2}$, this threshold is the same as that in (5) but improves on (5) if $d \geq 1/\sqrt{2}$. In particular, unlike the threshold in (5), the threshold in (15) is guaranteed to be at least as large as that in (7).

III. Robust Sparsity Thresholds

The deterministic sparsity thresholds for dictionaries in $\mathcal{D}(d,a,b)$ derived in the previous section (as those available in the literature for dictionaries in $\mathcal{D}_{\text{onb}}(d)$ and $\mathcal{D}_{\text{gen}}(d)$) all suffer from the so-called square-root bottleneck [11]. Specifically, from the Welch lower bound on coherence [21]

$$d \geq \sqrt{\frac{N-M}{M(N-1)}}$$

we can conclude that, for $N \gg M$, the deterministic sparsity thresholds reported in this paper scale as $\sqrt{M}$ as M grows large. Put differently, for a fixed number of nonzero entries S in $\mathbf{x}$, i.e., for a fixed sparsity level, the number of measurements M required to recover $\mathbf{x}$ through (P0), BP, or OMP is on

[6] Recall that for $d > 2(\sqrt{2} - 1)$ the threshold in (5) drops below that in (7).

the order of S^2. The square-root bottleneck stems from the fact that deterministic sparsity thresholds are universal thresholds in the sense of applying to all possible sparsity patterns (of cardinality S) and values of the corresponding nonzero entries of $\mathbf{x}$. As already mentioned in Section I, the probabilistic (i.e., robust) sparsity thresholds scale fundamentally better, namely according to $M/\log N$, which implies that the number of measurements required to recover $\mathbf{x}$ is on the order of $S \log N$ instead of S^2.

We next address the following question: Given a *general* dictionary, can we break the square-root bottleneck by only randomizing the sparsity patterns over a certain part of the overall dictionary? The answer turns out to be in the affirmative. It was shown in [17], [11]—for the concatenation of two ONBs—that randomization of the sparsity patterns is only required over one of the two ONBs. Before stating our results for general dictionaries let us briefly summarize the known results for concatenations of ONBs.

A. A Brief Review of Relevant Previous Work

Robust sparsity thresholds for the concatenation of two ONBs were first reported in [10] (based on earlier work in [17]) and later improved in [11]. In Theorem 5 below, we restate a result from [11] (obtained by combining Theorems D, 13, and 14) in a slightly modified form better suited to draw parallels to the case of dictionaries in $\mathcal{D}(d, a, b)$ considered in this paper.

Theorem 5 (Tropp, 2008): Assume that[7] $N > 2$. Let $\mathbf{D} \in \mathbb{C}^{M \times N}$ be the union of two ONBs for $\mathbb{C}^M$ given by $\mathbf{A}$ and $\mathbf{B}$ (i.e., $N = 2M$) and denote the coherence of $\mathbf{D}$ as d. Fix $s \geq 1$. Let the vector $\mathbf{x} \in \mathbb{C}^N$ have an *arbitrarily* chosen sparsity pattern of n_a nonzero entries corresponding to columns of sub-dictionary $\mathbf{A}$ and a *randomly* chosen sparsity pattern of n_b nonzero entries corresponding to columns of sub-dictionary $\mathbf{B}$. Suppose that

$$n_a + n_b < \min\left\{ \frac{c\, d^{-2}}{s \log N}, \frac{d^{-2}}{2} \right\} \tag{16}$$

where $c = 0.004212$. If the entries of $\mathbf{x}$ restricted to the chosen sparsity pattern are jointly continuous random variables,[8] then the unique solution of (P0) applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability exceeding $(1 - N^{-s})$.

If the total number of nonzero entries satisfies

$$n_a + n_b < \min\left\{ \frac{c\, d^{-2}}{s \log N}, \frac{d^{-2}}{2}, \frac{d^{-2}}{8(s+1)\log N} \right\} \tag{17}$$

[7] In [11] it is assumed that $M \geq 3$ (and hence $N \geq 6$). However, it can be shown that $N > 2$ is sufficient to establish the result.

[8] For a definition of joint continuity, we refer to [22, pp. 40].

and the entries of $\mathbf{x}$ restricted to the chosen sparsity pattern are jointly continuous random variables with i.i.d. phases that are uniformly distributed in $[0, 2\pi)$ (the magnitudes need not be i.i.d.), then the unique solution of both (P0) and BP applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability exceeding $(1 - 3N^{-s})$.

An important consequence of Theorem 5 is the following: For the concatenation of two ONBs a robust sparsity threshold $S = n_a + n_b$ of order $M/(\log N)$ is possible if the coherence d of the overall dictionary is on the order of $1/\sqrt{M}$. Note that for the same coherence d, deterministic sparsity thresholds would suffer from the square-root bottleneck as discussed in [11]. Remarkably, Theorem 5 does not require that the positions of *all* nonzero entries of $\mathbf{x}$ are chosen randomly: It suffices to pick the positions of the nonzero entries of $\mathbf{x}$ corresponding to one of the two ONBs at random, while the positions of the remaining nonzero entries—all corresponding to columns in the other ONB—can be chosen arbitrarily. This essentially means that the result is universal with respect to one of the two ONBs ($\mathbf{A}$ by choice of notation here) in the sense that *all* possible combinations of n_a columns in $\mathbf{A}$ are allowed. Randomization over the other ONB ensures that the overall sparsity patterns that cannot be recovered (with on the order of $S \log N$ measurements) are "weeded out". Moreover, randomization is needed on the values of *all* nonzero entries of $\mathbf{x}$, which reflects the fact that there exist certain value assignments on a given sparsity pattern that cannot be recovered with on the order of $S \log N$ measurements. In summary, Theorem 5 states that every sparsity pattern in $\mathbf{A}$ in conjunction with most sparsity patterns in $\mathbf{B}$ and most value assignments on the resulting overall sparsity pattern can be recovered.

This result is interesting as it hints at the possibility of isolating specific parts of the dictionary $\mathbf{D}$ that require randomization to "weed out" the support sets that are not recoverable. Unfortunately, the two-ONB structure is too restrictive to bring out this aspect. Specifically, as the two ONBs are on equal footing, the result in Theorem 5 does not allow us to understand which properties of a sub-dictionary are responsible for problematic sparsity patterns. This motivates looking at robust sparsity thresholds for the concatenation of two general dictionaries. Now, we could interpret the concatenation of two general (sub-)dictionaries as a general dictionary in $\mathcal{D}_{\text{gen}}(d)$ and apply the robust sparsity thresholds for general dictionaries reported in [11]. This requires, however, randomization over the entire dictionary (i.e., the positions of all nonzero entries of $\mathbf{x}$ have to be chosen at random and the values as well). Hence, the robust sparsity threshold for general dictionaries does not allow us to isolate specific parts of the dictionary $\mathbf{D}$ that require randomization to "weed out" the support sets that are not recoverable with on the order of $S \log N$ measurements.

B. Robust Sparsity Thresholds for the Concatenation of General Signal Sets

We next derive robust sparsity thresholds for dictionaries $\mathbf{D} \in \mathcal{D}(d, a, b)$. Our results not only generalize Theorem 5 to the concatenation of two general dictionaries but, since every dictionary in $\mathcal{D}_{\text{gen}}(d)$

can be viewed as the concatenation of two sub-dictionaries, also allow us to understand which part of a general dictionary requires randomization to “weed out” the support sets that are not recoverable.

Theorem 6: Assume that $N > 2$. Let $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ be a dictionary in $\mathcal{D}(d, a, b)$. Fix $s \geq 1$ and $\gamma \in [0, 1]$. Consider a random vector $\mathbf{x} = \begin{bmatrix}\mathbf{x}_a^T\ \mathbf{x}_b^T\end{bmatrix}^T$ where $\mathbf{x}_a \in \mathbb{C}^{N_a}$ has an *arbitrarily* chosen sparsity pattern of cardinality n_a such that

$$6\sqrt{2}\sqrt{n_a d^2 s \log N} + 2(n_a - 1)a \leq (1 - \gamma)e^{-1/4} \tag{18}$$

and $\mathbf{x}_b \in \mathbb{C}^{N_b}$ has a *randomly* chosen sparsity pattern of cardinality[9] n_b such that

$$24\sqrt{n_b b^2 s \log N} + \frac{4n_b}{N_b}\|\mathbf{B}\|^2 + 2\sqrt{\frac{n_b}{N_b}}\|\mathbf{A}\|\|\mathbf{B}\| \leq \gamma e^{-1/4}. \tag{19}$$

If the total number of nonzero entries of $\mathbf{x}$ satisfies

$$n_a + n_b \leq \frac{d^{-2}}{2} \tag{20}$$

and the entries of $\mathbf{x}$ restricted to the chosen sparsity pattern are jointly continuous random variables, then the unique solution of (P0) applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability exceeding $(1 - N^{-s})$.

If the total number of nonzero entries of $\mathbf{x}$ satisfies

$$n_a + n_b < \min\left\{\frac{d^{-2}}{2}, \frac{d^{-2}}{8(s+1)\log N}\right\} \tag{21}$$

and the entries of $\mathbf{x}$ restricted to the chosen sparsity pattern are jointly continuous random variables with i.i.d. phases that are uniformly distributed in $[0, 2\pi)$ (the magnitudes need not be i.i.d.), then the unique solution of both (P0) and BP applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability exceeding $(1 - 3N^{-s})$.

Proof: The proof is based on the following lemma proven in Appendix G.

Lemma 7: Fix $s \geq 1$ and $\gamma \in [0, 1]$. Let $\mathbf{S}$ be a sub-dictionary of $\mathbf{D} = [\mathbf{A}\ \mathbf{B}] \in \mathcal{D}(d, a, b)$ containing n_a *arbitrarily* chosen columns of $\mathbf{A}$ and n_b *randomly* chosen columns of $\mathbf{B}$. If n_a and n_b satisfy (18) and (19), respectively, then the minimum singular value $\sigma_{\min}(\mathbf{S})$ of the sub-dictionary $\mathbf{S}$ obeys

$$\mathbb{P}\left\{\sigma_{\min}(\mathbf{S}) \leq \frac{1}{\sqrt{2}}\right\} \leq N^{-s}.$$

The proof of Theorem 6 then follows from Lemma 7 and the results in [11] as follows. The sparsity pattern of $\mathbf{x}$ obtained according to the conditions in Theorem 6 induces a sub-dictionary $\mathbf{S}$ of $\mathbf{D}$ containing n_a arbitrarily chosen columns of $\mathbf{A}$ and n_b randomly chosen columns of $\mathbf{B}$. As a consequence

[9] Since we will be interested in the individual scaling behavior of n_a and n_b as M grows large, we shall assume in the remainder of the paper that $n_a, n_b \geq 1$.

of Lemma 7, the smallest singular value of this sub-dictionary exceeds $1/\sqrt{2}$ with probability at least $(1 - N^{-s})$.

Lemma 7 together with condition (20) and the requirement that the entries of $\mathbf{x}$ restricted to the chosen sparsity pattern are jointly continuous random variables implies, as a consequence of [11, Thm. 13] (see also Appendix H where [11, Thm. 13] is restated for completeness), that the unique solution of (P0) applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability at least $(1 - N^{-s})$.

The second statement in Theorem 6 is proven as follows. Lemma 7, together with condition (21), and the requirement that the entries of $\mathbf{x}$ restricted to the chosen sparsity pattern are jointly continuous random variables with i.i.d. phases that are uniformly distributed in $[0, 2\pi)$, implies, as a consequence of [11, Thm. 13] and [11, Thm. 14] (see also Appendix H), that the unique solution of *both* (P0) *and* BP applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability at least $(1 - N^{-s})(1 - 2N^{-s}) \geq (1 - 3N^{-s})$. ■

Theorem 6 generalizes the result in Theorem 5 to the concatenation $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ of the general dictionaries $\mathbf{A}$ and $\mathbf{B}$. Next, we determine conditions on $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ for breaking the square-root bottleneck. More precisely, we determine conditions on $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ such that for vectors $\mathbf{x}$ with[10] $n_a = \Theta(M/(\log N))$ and $n_b = \Theta(M/(\log N))$ the unique solution of *both* (P0) *and* BP applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability at least $1 - 3N^{-s}$. This implies a robust sparsity threshold $S = n_a + n_b$ of $\Theta(M/(\log N))$. Note that we say the square-root bottleneck is broken only if both n_a and n_b are on the order of $M/(\log N)$.

Conditions (18)–(21) in Theorem 6 yield upper bounds on the possible values of n_a and n_b (such that the unique solution of both (P0) and BP is given by $\mathbf{x}$) that depend on the dictionary parameters d, a, b, N_a, N_b, and the spectral norms of $\mathbf{A}$ and $\mathbf{B}$. In the following, we rewrite these upper bounds by absorbing all constants (including γ and s defined in Theorem 6) that are independent of d, a, b, N_a, N_b, and the spectral norms of $\mathbf{A}$ and $\mathbf{B}$ in a constant c. Note that c can take on a different value at each appearance. We then derive necessary and sufficient conditions on the dictionary parameters d, a, b, N_a, N_b, and the spectral norms of $\mathbf{A}$ and $\mathbf{B}$ for the resulting upper bounds on n_a and n_b to be on the order of $M/(\log N)$, respectively.

We start with condition (18), which, together with the obvious condition $n_a \leq N_a$, yields the following constraint on n_a:

$$n_a \leq c \min\left\{\frac{d^{-2}}{\log N}, a^{-1}, N_a\right\}.$$

[10]Whenever for some function $g(M, N)$ we write $\Theta(g(M, N))$, $\Omega(g(M, N))$, or $\mathcal{O}(g(M, N))$, we mean that the ratio M/N remains fixed while $M \to \infty$.

As M grows large, this upper bound is compatible[11] with the scaling behavior $n_a = \Theta(M/(\log N))$ if and only if all of the following conditions are met:

i) the coherence of $\mathbf{D}$ satisfies $d = \mathcal{O}(1/\sqrt{M})$

ii) the coherence of $\mathbf{A}$ satisfies $a = \mathcal{O}((\log N)/M)$

iii) the cardinality of $\mathbf{A}$ satisfies $N_a = \Omega(M/(\log N))$.

Similarly, we get from (19) that[12]

$$n_b \leq c \min\left\{ \frac{b^{-2}}{\log N}, \frac{N_b}{\|\mathbf{B}\|^2}, \frac{N_b}{\|\mathbf{A}\|^2 \, \|\mathbf{B}\|^2} \right\}. \tag{22}$$

This upper bound is compatible with the scaling behavior $n_b = \Theta(M/(\log N))$ if and only if all of the following conditions are met:

iv) the coherence of $\mathbf{B}$ satisfies $b = \mathcal{O}(1/\sqrt{M})$

v) the spectral norm of $\mathbf{B}$ satisfies $\|\mathbf{B}\|^2 \leq c\, N_b(\log N)/M$

vi) the spectral norm of $\mathbf{A}$ satisfies $\|\mathbf{A}\|^2 \leq c\, N_b(\log N)/(\|\mathbf{B}\|^2 M)$.

Note that iv) is implied by i) since $b \leq d$, by assumption. Finally, it follows from i) that conditions (20) and (21) are compatible with the scaling behavior $n_a = \Theta(M/(\log N))$ and $n_b = \Theta(M/(\log N))$.

In the special case of $\mathbf{A}$ and $\mathbf{B}$ being ONBs for $\mathbb{C}^M$, conditions ii) – vi) are trivially satisfied. Hence, in the two-ONB case the square-root bottleneck is broken by randomizing according to the specifications in Theorem 5 whenever $d = \mathcal{O}(1/\sqrt{M})$, as already shown in [11]. The additional requirements ii) – vi) become relevant for general dictionaries $\mathbf{D}$ only.

We next present an example of a non-trivial dictionary $\mathbf{D}$ with sub-dictionaries $\mathbf{A}$ and $\mathbf{B}$ (not both ONBs) that satisfy i) – vi). Let $M = p^k$, with p prime and $k \in \mathbb{N}^+$. For this choice of M it is possible to design $M+1$ ONBs for $\mathbb{C}^M$, which, upon concatenation, form a dictionary $\mathbf{D}$ with coherence d equal to $1/\sqrt{M}$ [23], [24], [8]. In particular, the absolute value of the inner product between two distinct columns of $\mathbf{D}$ is, by construction, either 0 or $1/\sqrt{M}$. Obviously, for such a dictionary i) is satisfied. Furthermore, identifying $\mathbf{A}$ with one of the $M+1$ ONBs and $\mathbf{B}$ with the concatenation of the remaining M ONBs, we have $a = 0$ and $N_a = M$. Hence ii) and iii) are satisfied. Since $\mathbf{B}$ consists of the concatenation of the remaining M ONBs, it has coherence $b = 1/\sqrt{M}$, and, hence, iv) is satisfied. Moreover, since $\mathbf{B}$ is the concatenation of M ONBs for $\mathbb{C}^M$, it forms a tight frame for $\mathbb{C}^M$. For a tight frame $\mathbf{B}$ with $N_b = M^2$ frame elements in $\mathbb{C}^M$ (all ℓ_2-normalized to one) the nonzero eigenvalues of the Gram matrix $\mathbf{B}^H\mathbf{B}$ are all equal to $N_b/M = M$. Hence, the spectral norm of $\mathbf{B}$ satisfies $\|\mathbf{B}\|^2 = M$. Thus, v) is met. Finally, since $\mathbf{A}$ is an ONB, its spectral norm satisfies $\|\mathbf{A}\|^2 = 1$ and, therefore, condition vi) is met. Now, as a

[11] We say that an upper bound on n_a, n_b is compatible with the scaling behavior $\Theta(M/(\log N))$, if it does not preclude this scaling behavior.

[12] Note that the obvious condition $n_b \leq N_b$ is implied by $n_b \leq N_b/\|\mathbf{B}\|^2$ since $\|\mathbf{B}\| \geq 1$.

consequence of Theorem 6, we obtain a robust sparsity threshold $S = n_a + n_b$ of order $M/(\log N)$ for the dictionary $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$. This threshold does not require that the positions of all nonzero entries of $\mathbf{x}$ are chosen randomly. Specifically, it suffices to randomize over the positions of the nonzero entries of $\mathbf{x}$ corresponding to $\mathbf{B}$, while the positions of the nonzero entries corresponding to $\mathbf{A}$ can be chosen arbitrarily. As for the two-ONB case, once the support set of $\mathbf{x}$ is chosen, the values of all nonzero entries of $\mathbf{x}$ need to be chosen at random.

Finally, as every dictionary $\mathbf{D} \in \mathcal{D}_{\text{gen}}(d)$ can be viewed as the concatenation of two general dictionaries $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$, we can now ask the following question: Given a general dictionary $\mathbf{D}$, over which part of the dictionary do we need to randomize to "weed out" the sparsity patterns that prohibit breaking the square-root bottleneck? From the results above we obtain the intuitive answer that in the "low-coherence" part of the dictionary, namely $\mathbf{A}$, we can pick the sparsity pattern arbitrarily, whereas the "high-coherence" part of the dictionary, namely $\mathbf{B}$, requires randomization. Note that, due to the bounds on the coherence parameters a and b in ii) and iv), respectively, the "low-coherence" part $\mathbf{A}$ of the overall dictionary $\mathbf{D}$ has, in general, fewer elements than the "high-coherence" part $\mathbf{B}$. Conditions i) – vi) can be used to identify the largest possible part $\mathbf{A}$ of the overall dictionary $\mathbf{D}$ where the corresponding sparsity pattern can be picked arbitrarily. Note, however, that the task of identifying the largest possible part $\mathbf{A}$ is in general difficult.

IV. Conclusion

We presented a generalization of the uncertainty relation for the representation of a signal in two different ONBs [2] to the representation of a signal in two general (possibly redundant or incomplete) signal sets. This novel uncertainty relation is important in the context of the analysis of signals containing two distinct features each of which can be described sparsely only in an overcomplete signal set. As shown in [25], the general uncertainty relation reported in this paper also forms the basis for establishing recovery guarantees for signals that are sparse in a (possibly overcomplete) dictionary and corrupted by noise that is also sparse in a (possibly overcomplete) dictionary.

We furthermore presented a novel deterministic sparsity threshold guaranteeing uniqueness of the (P0)-solution for general dictionaries $\mathbf{D} \in \mathcal{D}(d, a, b)$, as well as thresholds guaranteeing equivalence of this unique (P0)-solution to the solution obtained through BP and OMP. These thresholds improve on those previously known by up to a factor of two. Moreover, the known sparsity thresholds for general dictionaries and those for the concatenation of two ONBs follow from our results as special cases.

Finally, the probabilistic recovery guarantees presented in this paper allow us to understand which parts of a general dictionary one needs to randomize over to "weed out" the sparsity patterns that prohibit breaking the square-root bottleneck.

Appendix A
Proof of Lemma 1

Assume that $\mathbf{s} \in \mathbb{C}^M$ can be represented as a linear combination of n_a columns of $\mathbf{A}$ and, equivalently, as a linear combination of n_b columns of $\mathbf{B}$. This means that there exists a vector $\mathbf{p}$ with n_a nonzero entries and a vector $\mathbf{q}$ with n_b nonzero entries such that

$$\mathbf{s} = \mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}. \tag{23}$$

We exclude the trivial case $n_a = n_b = 0$ and note that for $n_a = 0$ or $n_b = 0$ we have $\mathbf{s} = \mathbf{0}_M$, by definition.

Left-multiplication in (23) by $\mathbf{A}^H$ yields

$$\mathbf{A}^H\mathbf{A}\mathbf{p} = \mathbf{A}^H\mathbf{B}\mathbf{q}. \tag{24}$$

We next lower-bound the absolute value of the ith entry ($i = 1, \ldots, N_a$) of the vector $\mathbf{A}^H\mathbf{A}\mathbf{p}$ according to

$$\begin{aligned}
\left|\left[\mathbf{A}^H\mathbf{A}\mathbf{p}\right]_i\right| &= \left|[\mathbf{p}]_i + \sum_{j\neq i}\left[\mathbf{A}^H\mathbf{A}\right]_{i,j}[\mathbf{p}]_j\right| \\
&\geq |[\mathbf{p}]_i| - a\sum_{j\neq i}|[\mathbf{p}]_j| \qquad (25)\\
&= (1+a)|[\mathbf{p}]_i| - a\|\mathbf{p}\|_1 \qquad (26)
\end{aligned}$$

where (25) follows from the reverse triangle inequality and the fact that the off-diagonal entries of $\mathbf{A}^H\mathbf{A}$ can be upper-bounded in absolute value by a. Next, we upper-bound the absolute value of the ith entry of the vector $\mathbf{A}^H\mathbf{B}\mathbf{q}$ as follows:

$$\left|\left[\mathbf{A}^H\mathbf{B}\mathbf{q}\right]_i\right| \leq d\|\mathbf{q}\|_1 . \tag{27}$$

Combining (26) and (27) yields

$$(1+a)|[\mathbf{p}]_i| - a\|\mathbf{p}\|_1 \leq d\|\mathbf{q}\|_1 .$$

If we now sum over all i for which $[\mathbf{p}]_i \neq 0$, we obtain

$$[(1+a) - n_a a]\|\mathbf{p}\|_1 \leq n_a d\|\mathbf{q}\|_1 \tag{28}$$

where we used that $\|\mathbf{p}\|_0 = n_a$, by assumption. Since $n_a d\|\mathbf{q}\|_1 \geq 0$, we can replace the LHS of (28) by the tighter bound

$$[(1+a) - n_a a]^+ \|\mathbf{p}\|_1 \leq n_a d\|\mathbf{q}\|_1 . \tag{29}$$

Multiplying both sides of (23) by $\mathbf{B}^H$ and following steps similar to the ones used to arrive at (29) yields

$$[(1+b)-n_b b]^+ \|\mathbf{q}\|_1 \leq n_b d \|\mathbf{p}\|_1 \,. \tag{30}$$

We now have to distinguish three cases. If both $n_a \geq 1$ and $n_b \geq 1$, and, hence, $\|\mathbf{p}\|_1 > 0$ and $\|\mathbf{q}\|_1 > 0$, we can combine (29) and (30) to obtain

$$n_a n_b d^2 \geq [(1+a)-n_a a]^+ \, [(1+b)-n_b b]^+ \,. \tag{31}$$

If $n_a = 0$ and $n_b \geq 1$ (i.e., $\|\mathbf{p}\|_1 = 0$ and $\|\mathbf{q}\|_1 > 0$), we get from (30) that

$$n_b \geq 1 + \frac{1}{b}. \tag{32}$$

Similarly, if $n_b = 0$ and $n_a \geq 1$ (i.e., $\|\mathbf{q}\|_1 = 0$ and $\|\mathbf{p}\|_1 > 0$), we obtain from (29) that

$$n_a \geq 1 + \frac{1}{a}. \tag{33}$$

Both (32) and (33) are contained in (31) as special cases, as is easily verified.

Appendix B
Proof of Theorem 2

The proof will be effected by deriving a lower bound on the spark of dictionaries in $\mathcal{D}(d,a,b)$, which together with (1), yields the desired result (10). This will be accomplished by finding a lower bound on the minimum number of nonzero entries that a nonzero vector $\mathbf{v} \in \mathbb{C}^N$ in the kernel of $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ must have. Without loss of generality, we may view $\mathbf{v}$ as the concatenation of two vectors $\mathbf{p} \in \mathbb{C}^{N_a}$ and $\mathbf{q} \in \mathbb{C}^{N_b}$, i.e., $\mathbf{v} = [\mathbf{p}^T\ \mathbf{q}^T]^T$. As $\mathbf{v}$ is in the kernel of $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$, we have

$$[\mathbf{A}\ \mathbf{B}] \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} = \mathbf{0}_M.$$

Therefore, the vectors $\mathbf{p}$ and $\mathbf{q}$ satisfy $\mathbf{A}\mathbf{p} = \mathbf{B}(-\mathbf{q}) \triangleq \mathbf{s}$. Let $n_a \triangleq \|\mathbf{p}\|_0$ and $n_b \triangleq \|-\mathbf{q}\|_0 = \|\mathbf{q}\|_0$ and recall that $n_a = 0$ is equivalent to $\mathbf{p} = \mathbf{0}_{N_a}$ and $n_b = 0$ is equivalent to $\mathbf{q} = \mathbf{0}_{N_b}$, both by definition. Since we require $\mathbf{v}$ to be a nonzero vector, the case of $n_a = n_b = 0$ (and hence $\mathbf{p} = \mathbf{0}_{N_a}$ and $\mathbf{q} = \mathbf{0}_{N_b}$, and, therefore $\mathbf{v} = \mathbf{0}_N$) is excluded. For all other cases, the uncertainty relation in Lemma 1 requires that the number of nonzero entries in $\mathbf{p}$ and $-\mathbf{q}$ (representing $\mathbf{s}$ according to $\mathbf{A}\mathbf{p} = \mathbf{B}(-\mathbf{q}) = \mathbf{s}$) satisfy

$$n_a n_b \geq \frac{[1-a(n_a-1)]^+ \, [1-b(n_b-1)]^+}{d^2}. \tag{34}$$

Based on (34), we now derive a lower bound on $\operatorname{spark}(\mathbf{D})$ by considering the following three different cases:

The case $n_b \geq 1$ and $n_a = 0$: In this case, the vector $\mathbf{v} = [\mathbf{p}^T \ \mathbf{q}^T]^T$ in the kernel of $\mathbf{D} = [\mathbf{A} \ \mathbf{B}]$ has nonzero entries only in the part $\mathbf{q}$ corresponding to sub-dictionary $\mathbf{B}$. It follows directly from (34) that

$$n_b \geq 1 + \frac{1}{b}. \tag{35}$$

The case $n_a \geq 1$ and $n_b = 0$: In this case, the vector $\mathbf{v} = [\mathbf{p}^T \ \mathbf{q}^T]^T$ in the kernel of $\mathbf{D} = [\mathbf{A} \ \mathbf{B}]$ has nonzero entries only in the part $\mathbf{p}$ corresponding to sub-dictionary $\mathbf{A}$. Again, direct application of (34) yields

$$n_a \geq 1 + \frac{1}{a}. \tag{36}$$

The case $n_a \geq 1$ and $n_b \geq 1$: In this case, the vector $\mathbf{v} = [\mathbf{p}^T \ \mathbf{q}^T]^T$ in the kernel of $\mathbf{D} = [\mathbf{A} \ \mathbf{B}]$ has nonzero entries in both parts $\mathbf{p}$ and $\mathbf{q}$ corresponding to sub-dictionary $\mathbf{A}$ and $\mathbf{B}$, respectively. Let $Z(\mathbf{D})$ denote the smallest possible number of nonzero entries of $\mathbf{v}$ in this case. Together with (35) and (36) we now have

$$\begin{aligned} \operatorname{spark}(\mathbf{D}) &\geq \min\left\{1 + \frac{1}{b}, 1 + \frac{1}{a}, Z(\mathbf{D})\right\} \\ &= \min\left\{1 + \frac{1}{b}, Z(\mathbf{D})\right\} \end{aligned} \tag{37}$$

where we used that $a \leq b$, by assumption. We next derive a lower bound on $Z(\mathbf{D})$ that is explicit in d, a, and b. Specifically, we minimize $n_a + n_b$ over all pairs (n_a, n_b) (with $n_a \geq 1$ and $n_b \geq 1$) that satisfy (34). Since, eventually, we are interested in finding a lower bound on $\operatorname{spark}(\mathbf{D})$, it follows from (37) that it suffices to restrict the minimization to those pairs (n_a, n_b), for which both $n_a \leq 1 + 1/b$ and $n_b \leq 1 + 1/b$. This implies that $[1 - a(n_a - 1)] \geq 0$ and $[1 - b(n_b - 1)] \geq 0$, and we thus have from (34) that

$$n_a n_b \geq \frac{[1 - a(n_a - 1)][1 - b(n_b - 1)]}{d^2}. \tag{38}$$

Solving (38) for n_a, we get

$$n_a \geq \frac{(1+a)(1+b) - n_b b(1+a)}{n_b(d^2 - ab) + a(1+b)} \triangleq f(n_b).$$

Finally, adding n_b on both sides yields

$$n_a + n_b \geq f(n_b) + n_b. \tag{39}$$

To arrive at a lower bound on $n_a + n_b$ that is explicit in d, a, and b only (in particular, the lower bound should be independent of n_a and n_b), we further lower-bound the RHS of (39) by minimizing $f(n_b) + n_b$ as a function of n_b, under the constraints $n_a \geq 1$ and $n_b \geq 1$ (implied by assumption). This yields the

following lower bound on $Z(\mathbf{D})$:[13]

$$Z(\mathbf{D}) \geq \min_{n_b \geq 1} [\max\{f(n_b), 1\} + n_b] \triangleq Z(d, a, b).$$

We now have that

$$\begin{aligned} Z(d, a, b) &= \min_{n_b \geq 1} [\max\{f(n_b), 1\} + n_b] \\ &\leq [\max\{f(n_b), 1\} + n_b]\big|_{n_b = 1/b} \\ &= 1 + \frac{1}{b} \end{aligned} \tag{40}$$

where we used the fact that $f(1/b) \leq 1$. As a consequence of (37), the inequality in (40) implies that

$$\begin{aligned} \operatorname{spark}(\mathbf{D}) &\geq Z(d, a, b) \\ &= \min_{n_b \geq 1} [\max\{f(n_b), 1\} + n_b] \\ &\geq \min_{x \geq 1} [\max\{f(x), 1\} + x] \end{aligned} \tag{41}$$

where (41) follows because minimizing over all $x \in \mathbb{R}$ with $x \geq 1$ yields a lower bound on the minimum taken over the integer parameter n_b only. We next compute the minimum in (41). The function $f(x)$ can be shown to be strictly decreasing. Furthermore, the equation $f(x) = 1$ has the unique solution $x_b = (1+b)/(b+d^2) \geq 1$, where the inequality follows because $d \leq 1$, by definition. We can therefore rewrite (41) as

$$\min_{x \geq 1} [\max\{f(x), 1\} + x] = \min_{1 \leq x \leq x_b} [f(x) + x]. \tag{42}$$

In the case $a = b = d$, the function $g(x) \triangleq f(x) + x$ reduces to the constant $1 + 1/d$ so that $\operatorname{spark}(\mathbf{D}) \geq 1 + 1/d$. In all other cases, the function $g(x)$ is strictly convex for $x \geq 0$. Furthermore, we have $g(1) \geq g(x_b)$ as a consequence of the assumption $a \leq b$. Hence, the minimum in (42) is attained either at the boundary point x_b, or at the stationary point x_s of $g(x)$, which is given by

$$x_s = \frac{d\sqrt{(1+a)(1+b)} - a(1+b)}{d^2 - ab} \geq 1. \tag{43}$$

The inequality in (43) follows from the convexity of $g(x)$ and the fact that $g(1) \geq g(x_b)$. If the stationary point x_s is inside the interval $[1, x_b]$, the minimum is attained at $\hat{x} = x_s$, otherwise it is attained at $\hat{x} = x_b$.

APPENDIX C
THE SPARSITY THRESHOLD IN THEOREM 2 IMPROVES ON THE THRESHOLD IN (7)

We show that the threshold in (10) improves on that in (7), unless $b = d$ or $d = 1$, in which case the threshold in (10) is the same as that in (7). This will be accomplished by considering the two (mutually

[13]The constraints $n_a \geq f(n_b)$ and $n_a \geq 1$ are combined into $n_a \geq \max\{f(n_b), 1\}$.

exclusive) cases $x_b \leq x_s$ and $x_b > x_s$.

The case $x_b \leq x_s$: The threshold in (10) equals

$$\frac{f(\hat{x}) + \hat{x}}{2} = \frac{f(x_b) + x_b}{2} = \frac{1}{2}\left(1 + \frac{1+b}{b+d^2}\right).$$

It is now easily verified that

$$\frac{1}{2}\left(1 + \frac{1+b}{b+d^2}\right) \geq \frac{1}{2}\left(1 + \frac{1}{d}\right)$$

for all $b \leq d \leq 1$ with equality if and only if $b = d$ or $d = 1$. Note that for $b = d$ (irrespective of a) or for $d = 1$ (irrespective of a and b), we have $x_b \leq x_s$.

The case $x_b > x_s$: Set $\Delta = \sqrt{(1+a)(1+b)} - d$. The function $f(x_s) + x_s$, which we denote as $h(a, b, d)$ to highlight its dependency on the variables a, b, and d, is strictly decreasing in a (for fixed b and d) as long as $b/d < \Delta < d/b$. Since $x_b > x_s$ implies $b < d$, and since $a \leq b$, by assumption, the inequality $\Delta < d/b$ is always satisfied. The inequality $b/d < \Delta$ holds whenever $x_s < x_b$, which is satisfied by assumption. Hence, we have that

$$\begin{aligned} \frac{f(x_s) + x_s}{2} = \frac{h(a,b,d)}{2} &\geq \frac{h(b,b,d)}{2} && (44) \\ &= \frac{1+b}{d+b} \\ &\geq \frac{1}{2}\left(1 + \frac{1}{d}\right). && (45) \end{aligned}$$

Note that equality in (44) and (45) holds if and only if $a = b = d$, already treated in the case $x_b \leq x_s$.

APPENDIX D
PROOF OF THEOREM 3

Our proof essentially follows the program laid out in [26] for dictionaries in $\mathcal{D}_{\text{onb}}(d)$, with appropriate modifications to account for the fact that we are dealing with the concatenation of two general dictionaries. Let $\mathbf{S}$ be the matrix that contains the columns of $\mathbf{A}$ and $\mathbf{B}$ participating in the representation of $\mathbf{y} = [\mathbf{A}\ \mathbf{B}]\,\mathbf{x}$, i.e., the columns in $[\mathbf{A}\ \mathbf{B}]$ corresponding to the nonzero entries in $\mathbf{x}$. A sufficient condition for BP *and* OMP applied to $\mathbf{y} = [\mathbf{A}\ \mathbf{B}]\,\mathbf{x}$ to recover $\mathbf{x}$ is [9, Thm. 3.1, Thm. 3.3]

$$\max_{\mathbf{d}_i} \left\| \mathbf{S}^\dagger \mathbf{d}_i \right\|_1 < 1 \tag{46}$$

where the maximization in (46) is performed over all columns $\mathbf{d}_i$ in $\mathbf{D}$ that do not appear in $\mathbf{S}$. We prove the theorem by first carefully bounding the absolute value of each element of the vector $\mathbf{S}^\dagger \mathbf{d}_i$. Concretely, we start with the following inequality

$$\left| [\mathbf{S}^\dagger \mathbf{d}_i]_k \right| = \left| [(\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{d}_i]_k \right|$$

$$= \left| \sum_l [(\mathbf{S}^H\mathbf{S})^{-1}]_{k,l} [\mathbf{S}^H \mathbf{d}_i]_l \right|$$
$$\leq \sum_l \left| [(\mathbf{S}^H\mathbf{S})^{-1}]_{k,l} \right| \left| [\mathbf{S}^H \mathbf{d}_i]_l \right|$$

and then bound the absolute value of each entry of the matrix $(\mathbf{S}^H\mathbf{S})^{-1}$ and of each element of the vector $\mathbf{S}^H\mathbf{d}_i$. We will verify below that the matrix $\mathbf{S}^H\mathbf{S}$ is invertible. To simplify notation, for any matrix $\mathbf{A}$, we let $|\mathbf{A}|$ be the matrix with entries

$$[|\mathbf{A}|]_{k,l} = |[\mathbf{A}]_{k,l}| \, .$$

Furthermore, if for two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size we have that

$$|[\mathbf{A}]_{k,l}| \leq |[\mathbf{B}]_{k,l}|$$

for all pairs (k, l), we shall write $|\mathbf{A}| \overset{\mathrm{e}}{\leq} |\mathbf{B}|$.

A. *Bound on the Elements of* $(\mathbf{S}^H\mathbf{S})^{-1}$

Since the columns of $\mathbf{D}$ are ℓ_2-normalized to 1, we can write

$$\mathbf{S}^H\mathbf{S} = \mathbf{I}_{n_a+n_b} - \mathbf{K}$$

where $-\mathbf{K}$ contains the off-diagonal elements of $\mathbf{S}^H\mathbf{S}$. Clearly,

$$|\mathbf{K}| \overset{\mathrm{e}}{\leq} \begin{bmatrix} a(\mathbf{1}_{n_a,n_a} - \mathbf{I}_{n_a}) & d\mathbf{1}_{n_a,n_b} \\ d\mathbf{1}_{n_b,n_a} & b(\mathbf{1}_{n_b,n_b} - \mathbf{I}_{n_b}) \end{bmatrix} \overset{\mathrm{e}}{\leq} \begin{bmatrix} b(\mathbf{1}_{n_a,n_a} - \mathbf{I}_{n_a}) & d\mathbf{1}_{n_a,n_b} \\ d\mathbf{1}_{n_b,n_a} & b(\mathbf{1}_{n_b,n_b} - \mathbf{I}_{n_b}) \end{bmatrix}$$
$$= -b\mathbf{I}_{n_a+n_b} + d\mathbf{1}_{n_a+n_b,n_a+n_b} - (d-b)\mathbf{T} \tag{47}$$

where we set

$$\mathbf{T} = \begin{bmatrix} \mathbf{1}_{n_a,n_a} & \mathbf{0}_{n_a,n_b} \\ \mathbf{0}_{n_b,n_a} & \mathbf{1}_{n_b,n_b} \end{bmatrix}.$$

As a consequence of (47) and using the assumption $n_b \geq n_a$, we have that $\|\mathbf{K}\|_{1,1} \leq dn_b + b(n_a - 1)$. Since $\|\cdot\|_{1,1}$ is a matrix norm [27, p. 294], the matrix $\mathbf{S}^H\mathbf{S}$ is invertible whenever $dn_b + b(n_a - 1) < 1$, and, moreover, we can expand $(\mathbf{S}^H\mathbf{S})^{-1}$ into a Neumann series according to $(\mathbf{S}^H\mathbf{S})^{-1} = \mathbf{I}_{n_a+n_b} + \sum_{k=1}^{\infty} \mathbf{K}^k$. As the condition in (12) implies that $dn_b + b(n_a - 1) < 1$, we have

$$\left|(\mathbf{S}^H\mathbf{S})^{-1}\right| = \left| \mathbf{I}_{n_a+n_b} + \sum_{k=1}^{\infty} \mathbf{K}^k \right| \tag{48}$$
$$\overset{\mathrm{e}}{\leq} \mathbf{I}_{n_a+n_b} + \sum_{k=1}^{\infty} |\mathbf{K}|^k \tag{49}$$

$$\overset{\text{e}}{\leq} \mathbf{I}_{n_a+n_b} + \sum_{k=1}^{\infty} [-b\mathbf{I}_{n_a+n_b} + d\mathbf{1}_{n_a+n_b,n_a+n_b} - (d-b)\mathbf{T}]^k$$
$$= \big[\underbrace{(1+b)\mathbf{I}_{n_a+n_b} + (d-b)\mathbf{T}}_{\triangleq \mathbf{X}} - d\mathbf{1}_{n_a+n_b,n_a+n_b}\big]^{-1}$$
$$= \left[\mathbf{I}_{n_a+n_b} - d\mathbf{X}^{-1}\mathbf{1}_{n_a+n_b,n_a+n_b}\right]^{-1}\mathbf{X}^{-1}. \tag{50}$$

Here, in (49) we used the triangle inequality and the fact that $\left|\mathbf{K}^k\right| \overset{\text{e}}{\leq} |\mathbf{K}|^k$. We next compute the inverses in (50). To get $\mathbf{X}^{-1}$, we use the fact that $\mathbf{X}$ is a block-diagonal matrix and apply Woodbury's identity [27, p.19] to each of the two blocks,[14] which yields

$$\mathbf{X}^{-1} = \begin{bmatrix} \frac{1}{1+b}\left(\mathbf{I}_{n_a} - \frac{d-b}{(d-b)n_a+1+b}\mathbf{1}_{n_a,n_a}\right) & \mathbf{0}_{n_a,n_b} \\ \mathbf{0}_{n_b,n_a} & \frac{1}{1+b}\left(\mathbf{I}_{n_b} - \frac{d-b}{(d-b)n_b+1+b}\mathbf{1}_{n_b,n_b}\right) \end{bmatrix}. \tag{51}$$

Next, setting $c_a = [(d-b)n_a + 1 + b]^{-1}$, $c_b = [(d-b)n_b + 1 + b]^{-1}$, and

$$\mathbf{v} = d \begin{bmatrix} c_a \mathbf{1}_{n_a} \\ c_b \mathbf{1}_{n_b} \end{bmatrix}$$

steps similar to the ones reported in [26, Eq. (A.2)-(A.3)] yield

$$\left[\mathbf{I}_{n_a+n_b} - d\mathbf{X}^{-1}\mathbf{1}_{n_a+n_b,n_a+n_b}\right]^{-1} = \mathbf{I}_{n_a+n_b} + \frac{1}{1-d(c_a n_a + c_b n_b)}\mathbf{v}\mathbf{1}^T_{n_a+n_b}. \tag{52}$$

Using the fact, shown in (50), that

$$\left|(\mathbf{S}^H\mathbf{S})^{-1}\right| \overset{\text{e}}{\leq} \left[\mathbf{I}_{n_a+n_b} - d\mathbf{X}^{-1}\mathbf{1}_{n_a+n_b,n_a+n_b}\right]^{-1}\mathbf{X}^{-1}$$

we can combine (51) and (52) to obtain an upper bound on the absolute value of each entry of $(\mathbf{S}^H\mathbf{S})^{-1}$.

B. Bound on the Elements of $\mathbf{S}^H\mathbf{d}_i$

Let $\mathbf{d}_i$ be a column of $\mathbf{D}$ that does not appear in $\mathbf{S}$. Assume that $\mathbf{d}_i \in \mathbf{A}$ (we will later show that in searching the maximum in (46) it is, indeed, sufficient to assume $\mathbf{d}_i \in \mathbf{A}$). Then, we have

$$\left|\mathbf{S}^H\mathbf{d}_i\right| \overset{\text{e}}{\leq} \begin{bmatrix} a\mathbf{1}_{n_a} \\ d\mathbf{1}_{n_b} \end{bmatrix} \overset{\text{e}}{\leq} \begin{bmatrix} b\mathbf{1}_{n_a} \\ d\mathbf{1}_{n_b} \end{bmatrix}. \tag{53}$$

As a sideremark, we note that we loose the dependency of our final result on a through the bounds (47) and (53).

[14]To apply Woodbury's identity, we exploit the fact that $\mathbf{1}_{n,n} = \mathbf{1}_n\mathbf{1}_n^T$.

C. Putting the Pieces Together

Substituting (52) into (50), we get

$$\begin{aligned}\left|\mathbf{S}^{\dagger}\mathbf{d}_i\right| &\overset{\mathrm{e}}{\leq} \left(\mathbf{I}_{n_a+n_b} + \frac{1}{1-d(c_a n_a + c_b n_b)}\mathbf{v}\mathbf{1}^T_{n_a+n_b}\right)\mathbf{X}^{-1}\begin{bmatrix} b\mathbf{1}_{n_a} \\ d\mathbf{1}_{n_b}\end{bmatrix} \\ &= \left(\mathbf{I}_{n_a+n_b} + \frac{1}{1-d(c_a n_a + c_b n_b)}\mathbf{v}\mathbf{1}^T_{n_a+n_b}\right)\begin{bmatrix} bc_a\mathbf{1}_{n_a} \\ dc_b\mathbf{1}_{n_b}\end{bmatrix} \\ &= \frac{1}{1-d(c_a n_a + c_b n_b)}\begin{bmatrix}(bc_a + (d-b)dn_b c_a c_b)\mathbf{1}_{n_a} \\ (dc_b - (d-b)dn_a c_a c_b)\mathbf{1}_{n_b}\end{bmatrix}. \end{aligned} \tag{54}$$

Summing the RHS of (54) over all entries of the vector $\mathbf{S}^{\dagger}\mathbf{d}_i$ yields the following upper bound on $\left\|\mathbf{S}^{\dagger}\mathbf{d}_i\right\|_1$:

$$\left\|\mathbf{S}^{\dagger}\mathbf{d}_i\right\|_1 \leq \frac{bc_a n_a + dc_b n_b}{1-d(c_a n_a + c_b n_b)}. \tag{55}$$

If we instead assume that $\mathbf{d}_i \in \mathbf{B}$ and apply the same steps as before, we find that

$$\left\|\mathbf{S}^{\dagger}\mathbf{d}_i\right\|_1 \leq \frac{dc_a n_a + bc_b n_b}{1-d(c_a n_a + c_b n_b)}. \tag{56}$$

Since $bc_a n_a + dc_b n_b \geq dc_a n_a + bc_b n_b$ it follows that

$$\frac{dc_a n_a + bc_b n_b}{1-d(c_a n_a + c_b n_b)} \leq \frac{bc_a n_a + dc_b n_b}{1-d(c_a n_a + c_b n_b)}$$

and hence

$$\max_{\mathbf{d}_i}\left\|\mathbf{S}^{\dagger}\mathbf{d}_i\right\|_1 \leq \frac{bc_a n_a + dc_b n_b}{1-d(c_a n_a + c_b n_b)}.$$

We can therefore conclude that a sufficient condition for BP and OMP applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ to recover $\mathbf{x}$ is

$$\frac{bc_a n_a + dc_b n_b}{1-d(c_a n_a + c_b n_b)} < 1. \tag{57}$$

Simple algebraic manipulations reveal that (57) is equivalent to (12).

APPENDIX E
PROOF OF COROLLARY 4

We obtain Corollary 4 as a consequence of Theorem 3 as follows. For given $n_b \geq n_a$ it follows from (12) that a sufficient condition for BP and OMP to recover the unknown vector $\mathbf{x}$ is

$$n_a < \frac{(1+b)^2 - n_b(1+b)(d+b)}{2b(1+b) + 2n_b(d^2-b^2)} \triangleq h(n_b).$$

To arrive at a sparsity threshold that is explicit in b and d only, we minimize $h(n_b) + n_b$ over n_b, under the constraint $n_b \geq 1$ (recall that $n_b \geq n_a$ and note that representing a nonzero vector $\mathbf{y} \in \mathbb{C}^M$ requires

at least one column of $\mathbf{D}$). Furthermore, we have that

$$\min_{n_b \geq 1}[h(n_b) + n_b] \geq \min_{x \geq 1}[h(x) + x] \triangleq S \tag{58}$$

where $x \in \mathbb{R}$. Clearly, minimizing over all $x \geq 1$ with $x \in \mathbb{R}$, as opposed to integer values n_b only, can only yield a smaller value for the minimum. In the case $b = d$, the function $h(x) + x$ reduces to the constant $(1 + 1/d)/2$, thereby recovering the previously known sparsity threshold in (7). In all other cases, the function $h(x) + x$ is strictly convex for $x \geq 0$. Hence, the minimum in (58) is attained either at the boundary point $x = 1$ or at the stationary point x_s of $h(x) + x$, given by

$$x_s = \frac{(1+b)(\sqrt{2d(b+d)} - 2b)}{2(d^2 - b^2)}.$$

If the stationary point satisfies $x_s > 1$, then the minimum in (58) is attained at the stationary point, otherwise the minimum is attained at the boundary point $x = 1$. The condition $x_s > 1$ is equivalent to the condition $\kappa(d, b) > 1$ (where $\kappa(d, b)$ is defined in (14)). If $\kappa(d, b) > 1$ the minimum in (58) is given by

$$S = \frac{(1+b)\left[2\sqrt{2}\sqrt{d(b+d)} - (d+3b)\right]}{2(d^2 - b^2)}.$$

If $\kappa(d, b) \leq 1$, the minimum in (58) is attained at the boundary point $x = 1$ and is given by

$$S = \frac{1 + 2d^2 + 3b - d(1+b)}{2(d^2 + b)}. \tag{59}$$

Note that for $b = d$ the sparsity threshold in (59) reduces to that in (7).

Appendix F

The Sparsity Threshold in Corollary 4 Improves on the Threshold in (7)

We show that the threshold in Corollary 4 improves on that in (7), unless $b = d$ or $d = 1$, in which case the threshold in Corollary 4 is the same as that in (7). Let us first consider the case when the RHS of (13) in Corollary 4 reduces to

$$S \triangleq \frac{1 + 2d^2 + 3b - d(1+b)}{2(d^2 + b)}.$$

We need to establish that

$$\frac{1 + 2d^2 + 3b - d(1+b)}{2(d^2 + b)} \geq \frac{1}{2}\left(1 + \frac{1}{d}\right) \tag{60}$$

with equality if and only if $b = d$ or $d = 1$. Straightforward calculations reveal that the inequality (60) is equivalent to

$$(d - b)(1 - d)^2 \geq 0 \tag{61}$$

which is satisfied for all $b \leq d$. Furthermore, equality in (61) holds if and only if $b = d$ or $d = 1$.

Next, we consider the case $b < d$ and $\kappa(d,b) > 1$ so that the RHS of (13) reduces to

$$S = \frac{(1+b)\left[2\sqrt{2}\sqrt{d(b+d)} - (d+3b)\right]}{2(d^2-b^2)}.$$

For $d \leq 7/9$ it can be verified that

$$\frac{(1+b)\left[2\sqrt{2}\sqrt{d(b+d)} - (d+3b)\right]}{2(d^2-b^2)} > \frac{1}{2}\left(1+\frac{1}{d}\right).$$

It turns out that a necessary condition for $\kappa(d,b) > 1$ is $d < 1/\sqrt{2}$. The proof is completed by noting that $1/\sqrt{2} < 7/9$.

Appendix G
Proof of Lemma 7

Since the minimum singular value $\sigma_{\min}(\mathbf{S})$ of the sub-dictionary $\mathbf{S}$ can be lower-bounded as $\sigma^2_{\min}(\mathbf{S}) \geq 1 - \left\|\mathbf{S}^H\mathbf{S} - \mathbf{I}_{n_a+n_b}\right\|$, we have

$$\begin{aligned}
\mathbb{P}\left\{\sigma_{\min}(\mathbf{S}) \leq \frac{1}{\sqrt{2}}\right\} &= \mathbb{P}\left\{\sigma^2_{\min}(\mathbf{S}) \leq \frac{1}{2}\right\} \\
&\leq \mathbb{P}\left\{1 - \left\|\mathbf{S}^H\mathbf{S} - \mathbf{I}_{n_a+n_b}\right\| \leq \frac{1}{2}\right\} \\
&= \mathbb{P}\left\{\left\|\mathbf{S}^H\mathbf{S} - \mathbf{I}_{n_a+n_b}\right\| \geq \frac{1}{2}\right\}. \qquad (62)
\end{aligned}$$

Next, we study the tail behavior of the random variable $H = \left\|\mathbf{S}^H\mathbf{S} - \mathbf{I}_{n_a+n_b}\right\|$, which will then allow us to upper-bound $\mathbb{P}\left\{\left\|\mathbf{S}^H\mathbf{S} - \mathbf{I}_{n_a+n_b}\right\| \geq 1/2\right\}$. To this end the following lemma, which follows from Markov's inequality, will be useful.

Lemma 8 ([11, Prop. 10]): If the moments of a nonnegative random variable R can be upper-bounded as $[\mathbb{E}(R^q)]^{1/q} \leq \alpha\sqrt{q} + \beta$ for all $q \geq Q \geq 1$, where $\alpha, \beta > 0$, then,

$$\mathbb{P}\{R \geq e^{1/4}(\alpha u + \beta)\} \leq e^{-u^2/4}$$

for all $u \geq \sqrt{Q}$.

To be able to apply Lemma 8 to $H = \left\|\mathbf{S}^H\mathbf{S} - \mathbf{I}_{n_a+n_b}\right\|$, we first need an upper bound on $[\mathbb{E}(H^q)]^{1/q}$ that is of the form $\alpha\sqrt{q} + \beta$. To derive this upper bound, we start by writing $\mathbf{S}$ as $\mathbf{S} = [\mathbf{S}_a\ \mathbf{S}_b]$, where $\mathbf{S}_a$ and $\mathbf{S}_b$ denote the matrices containing the columns chosen arbitrarily from $\mathbf{A}$ and randomly from $\mathbf{B}$, respectively. We then obtain

$$\mathbf{S}^H\mathbf{S} - \mathbf{I}_{n_a+n_b} = \begin{bmatrix} \mathbf{S}_a^H\mathbf{S}_a - \mathbf{I}_{n_a} & \mathbf{S}_a^H\mathbf{S}_b \\ \mathbf{S}_b^H\mathbf{S}_a & \mathbf{S}_b^H\mathbf{S}_b - \mathbf{I}_{n_b} \end{bmatrix}.$$

Applying the triangle inequality for operator norms, we can now upper-bound H according to

$$\begin{aligned}
H &= \left\| \begin{bmatrix} \mathbf{S}_a^H \mathbf{S}_a - \mathbf{I}_{n_a} & \mathbf{S}_a^H \mathbf{S}_b \\ \mathbf{S}_b^H \mathbf{S}_a & \mathbf{S}_b^H \mathbf{S}_b - \mathbf{I}_{n_b} \end{bmatrix} \right\| \\
&\leq \left\| \begin{bmatrix} \mathbf{S}_a^H \mathbf{S}_a - \mathbf{I}_{n_a} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_b^H \mathbf{S}_b - \mathbf{I}_{n_b} \end{bmatrix} \right\| + \left\| \begin{bmatrix} \mathbf{0} & \mathbf{S}_a^H \mathbf{S}_b \\ \mathbf{S}_b^H \mathbf{S}_a & \mathbf{0} \end{bmatrix} \right\| \\
&\leq \max\left\{ \left\| \mathbf{S}_a^H \mathbf{S}_a - \mathbf{I}_{n_a} \right\|, \left\| \mathbf{S}_b^H \mathbf{S}_b - \mathbf{I}_{n_b} \right\| \right\} + \left\| \mathbf{S}_a^H \mathbf{S}_b \right\| \\
&\leq \left\| \mathbf{S}_a^H \mathbf{S}_a - \mathbf{I}_{n_a} \right\| + \left\| \mathbf{S}_b^H \mathbf{S}_b - \mathbf{I}_{n_b} \right\| + \left\| \mathbf{S}_a^H \mathbf{S}_b \right\| \qquad (63)
\end{aligned}$$

where the second inequality follows because the spectral norm of both a block-diagonal matrix and an anti-block-diagonal matrix is given by the largest among the spectral norms of the individual nonzero blocks. Next, we define $H_a = \left\| \mathbf{S}_a^H \mathbf{S}_a - \mathbf{I}_{n_a} \right\|$, $H_b = \left\| \mathbf{S}_b^H \mathbf{S}_b - \mathbf{I}_{n_b} \right\|$, and $Z = \left\| \mathbf{S}_a^H \mathbf{S}_b \right\|$. It then follows from (63) that for all $q \geq 1$

$$\begin{aligned}
\left[\mathbb{E}(H^q)\right]^{1/q} &\leq \left[\mathbb{E}((H_a + H_b + Z)^q)\right]^{1/q} \\
&\leq \left[\mathbb{E}(H_a^q)\right]^{1/q} + \left[\mathbb{E}(H_b^q)\right]^{1/q} + \left[\mathbb{E}(Z^q)\right]^{1/q} \\
&= H_a + \left[\mathbb{E}(H_b^q)\right]^{1/q} + \left[\mathbb{E}(Z^q)\right]^{1/q} \qquad (64)
\end{aligned}$$

where the second inequality is a consequence of the triangle inequality for the norm $[\mathbb{E}(|\cdot|^q)]^{1/q}$ (recall that we assumed $q \geq 1$ and hence $[\mathbb{E}(|\cdot|^q)]^{1/q}$ is a norm), and in the last step we used the fact that H_a is a deterministic quantity. All expectations in (64) are with respect to the random choice of columns from the sub-dictionary $\mathbf{B}$.

We next upper-bound the three terms on the RHS of (64) individually. Applying Geršgorin's disc theorem [27, Thm. 6.1.1] to the first term, we obtain

$$H_a = \left\| \mathbf{S}_a^H \mathbf{S}_a - \mathbf{I}_{n_a} \right\| \leq (n_a - 1)a. \qquad (65)$$

For the second term, we use [11, Eq. (6.1)] to get

$$\begin{aligned}
\left[\mathbb{E}(H_b^q)\right]^{1/q} &= \left[\mathbb{E}\left(\left\| \mathbf{S}_b^H \mathbf{S}_b - \mathbf{I}_{n_b} \right\|^q \right)\right]^{1/q} \\
&\leq \sqrt{144 b^2 n_b r_1} + \frac{2 n_b}{N_b} \|\mathbf{B}\|^2 \qquad (66)
\end{aligned}$$

where $r_1 = \max\{1, \log(n_b/2 + 1), q/4\}$. Assuming that $q \geq \max\{4 \log(n_b/2 + 1), 4\}$ and, hence, $r_1 = q/4$, we can simplify (66) to

$$\left[\mathbb{E}(H_b^q)\right]^{1/q} \leq 6 \sqrt{b^2 n_b} \sqrt{q} + \frac{2 n_b}{N_b} \|\mathbf{B}\|^2 . \qquad (67)$$

To bound the third term, we use the upper bound in [11, Thm. 8] on the spectral norm of a random compression combined with the fact that $\mathrm{rank}(\mathbf{S}_a^H\mathbf{S}_b) \leq n_b$, which is a consequence of $\mathbf{S}_a^H\mathbf{S}_b$ being of dimension $n_a \times n_b$. This yields

$$\begin{aligned}[\mathbb{E}(Z^q)]^{1/q} &= \left[\mathbb{E}\Big(\left\|\mathbf{S}_a^H\mathbf{S}_b\right\|^q\Big)\right]^{1/q} \\ &\leq 3\sqrt{r_2}\left\|\mathbf{S}_a^H\mathbf{B}\right\|_{1,2} + \sqrt{\frac{n_b}{N_b}}\left\|\mathbf{S}_a^H\mathbf{B}\right\| \qquad (68)\end{aligned}$$

where $r_2 = \max\{2, 2\log n_b, q/2\}$. Assuming that $q \geq \max\{4\log n_b, 4\}$, we can further upper-bound the RHS of (68) to get

$$\begin{aligned}[\mathbb{E}(Z^q)]^{1/q} &\leq \frac{3}{\sqrt{2}}\sqrt{q}\left\|\mathbf{S}_a^H\mathbf{B}\right\|_{1,2} + \sqrt{\frac{n_b}{N_b}}\left\|\mathbf{S}_a^H\mathbf{B}\right\| \\ &\leq \frac{3}{\sqrt{2}}\sqrt{d^2n_a}\sqrt{q} + \sqrt{\frac{n_b}{N_b}}\left\|\mathbf{S}_a^H\mathbf{B}\right\| \qquad (69) \\ &\leq \frac{3}{\sqrt{2}}\sqrt{d^2n_a}\sqrt{q} + \sqrt{\frac{n_b}{N_b}}\|\mathbf{A}\|\,\|\mathbf{B}\| \qquad (70)\end{aligned}$$

where (69) follows from the fact that the magnitude of each entry of $\mathbf{S}_a^H\mathbf{B}$ is upper-bounded by d and, thus, $\left\|\mathbf{S}_a^H\mathbf{B}\right\|_{1,2} \leq \sqrt{d^2n_a}$. To arrive at (70) we used $\left\|\mathbf{S}_a^H\mathbf{B}\right\| \leq \left\|\mathbf{S}_a^H\right\|\|\mathbf{B}\| \leq \|\mathbf{A}\|\,\|\mathbf{B}\|$, which follows from the sub-multiplicativity of the spectral norm and the fact that the spectral norm of the submatrix $\mathbf{S}_a$ of $\mathbf{A}$ cannot exceed that of $\mathbf{A}$ [27, Thm. 4.3.3]. We can now combine the upper bounds (65), (67), and (70) to obtain

$$\begin{aligned}[\mathbb{E}(H^q)]^{1/q} &\leq (n_a-1)a + 6\sqrt{b^2n_b}\sqrt{q} + \frac{2n_b}{N_b}\|\mathbf{B}\|^2 + \\ &\quad + \frac{3}{\sqrt{2}}\sqrt{d^2n_a}\sqrt{q} + \sqrt{\frac{n_b}{N_b}}\|\mathbf{A}\|\,\|\mathbf{B}\| \\ &= \underbrace{\left(6\sqrt{b^2n_b} + \frac{3}{\sqrt{2}}\sqrt{d^2n_a}\right)}_{\alpha}\sqrt{q} + \\ &\quad \underbrace{+\,(n_a-1)a + \frac{2n_b}{N_b}\|\mathbf{B}\|^2 + \sqrt{\frac{n_b}{N_b}}\|\mathbf{A}\|\,\|\mathbf{B}\|}_{\beta} \\ &= \alpha\sqrt{q} + \beta\end{aligned}$$

for all $q \geq Q_1 = \max\{4\log(n_b/2+1), 4\log n_b, 4\}$. Hence, Lemma 8 yields

$$\mathbb{P}\{H \geq e^{1/4}(\alpha u + \beta)\} \leq e^{-u^2/4}$$

for all $u \geq \sqrt{Q_1}$. In particular, under the assumption $N \geq e \approx 2.7$, it follows that the choice $u = \sqrt{4s\log N}$ satisfies $u \geq \sqrt{Q_1}$ for $s \geq 1$. Straightforward calculations reveal that conditions (18) and (19)

ensure that $e^{1/4}(\alpha u + \beta) \leq 1/2$, which together with (62) leads to

$$\begin{aligned}\mathbb{P}\left\{\sigma_{\min}(\mathbf{S}) \leq 1/\sqrt{2}\right\} &\leq \mathbb{P}\{H \geq 1/2\} \\ &\leq \mathbb{P}\{H \geq e^{1/4}(\alpha u + \beta)\} \\ &\leq e^{-u^2/4} = N^{-s}.\end{aligned}$$

APPENDIX H
PRIOR ART

A. *Tropp's (M0) Model and (P0)-uniqueness*

In [11] the following model was introduced.

Model (M0) for a signal $\mathbf{y} = \mathbf{D}\mathbf{x}$		
The dictionary	$\mathbf{D}$	has coherence d.
The vector	$\mathbf{x}$	has nonzero entries only in the positions corresponding to the columns of a sub-dictionary $\mathbf{S}$ of $\mathbf{D}$; furthermore, the entries of $\mathbf{x}$ restricted to the chosen sparsity pattern are jointly continuous random variables.
The sub-dictionary	$\mathbf{S}$	satisfies $\sigma_{\min}(\mathbf{S}) \geq 1/\sqrt{2}$ and has $T < d^{-2}/2$ columns.

The following theorem builds on (M0).

Theorem 9 ([11, Thm. 13]): Suppose that $\mathbf{y} = \mathbf{D}\mathbf{x}$ is a signal drawn from Model (M0). Then $\mathbf{x}$ is almost surely the unique vector that satisfies the constraints

$$\mathbf{D}\mathbf{x} = \mathbf{y} \quad \text{and} \quad \|\mathbf{x}\|_0 \leq T.$$

B. *Tropp's (M1) Model and Recovery via BP*

In [11] the following model was introduced.

Model (M1) for a signal $\mathbf{y} = \mathbf{D}\mathbf{x}$		
The dictionary	$\mathbf{D}$	has coherence d.
The vector	$\mathbf{x}$	has nonzero entries only in the positions corresponding to the columns of a sub-dictionary $\mathbf{S}$ of $\mathbf{D}$; furthermore, the phases of its nonzero entries are i.i.d. and uniformly distributed on $[0, 2\pi)$ (the magnitudes need not be i.i.d.).
The sub-dictionary	$\mathbf{S}$	satisfies $\sigma_{\min}(\mathbf{S}) \geq 1/\sqrt{2}$ and has $T < d^{-2}/[8(s+1)\log N]$ columns ($s \geq 1$).

The following theorem builds on (M1).

Theorem 10 ([11, Thm. 14]): Suppose that $\mathbf{y} = \mathbf{D}\mathbf{x}$ is a signal drawn from Model (M1). Then $\mathbf{x}$ is the unique solution of (BP) with probability at least $1 - 2N^{-s}$.

If the requirements of both (M0) and (M1) are satisfied, then combining Theorems 9 and 10 yields the following statement: The unique solution of *both* (P0) and BP applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ is given by $\mathbf{x}$ with probability at least $1 - 2N^{-s}$. Note, however, that both (M0) and (M1) require the sub-dictionary $\mathbf{S}$ to have $\sigma_{\min}(\mathbf{S}) \geq 1/\sqrt{2}$. Lemma 7 shows that for $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ and $\mathbf{S}$ consisting of n_a arbitrarily chosen columns of $\mathbf{A}$ and n_b randomly chosen columns of $\mathbf{B}$ the sub-dictionary $\mathbf{S}$ has $\sigma_{\min}(\mathbf{S}) \geq 1/\sqrt{2}$ with probability at least $1 - N^{-s}$.

REFERENCES

[1] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math*, vol. 49, no. 3, pp. 906–931, Jun. 1989.

[2] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[3] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA. U.S.A.: Birkhäuser, 2003.

[4] M. Fadili, J. Starck, J. Bobin, and Y. Moudden, "Image decomposition and separation using sparse representations: An overview," *Proc. of the IEEE*, vol. 98, no. 6, pp. 983–994, Jun. 2010.

[5] D. L. Donoho and G. Kutyniok, "Microlocal analysis of the geometric separation problem," Preprint: arXiv:1004.3006v1 [math.FA], Feb. 2010.

[6] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[8] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[9] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[10] E. J. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Comput. Math.*, vol. 6, no. 2, pp. 227–254, Apr. 2006.

[11] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 1–24, Jul. 2008.

[12] G. Davis, S. G. Mallat, and M. Avellaneda, "Adaptive greedy algorithms," *Constr. Approx.*, vol. 13, pp. 57–98, 1997.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[14] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Proc. of 27th Asilomar Conf. on Signals, Systems, and Comput., Pacific Grove, CA, USA*, pp. 40–44, Nov. 1993.

[15] G. Davis, S. G. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Opt. Eng.*, vol. 33, no. 7, pp. 2183–2191, Jul. 1994.

[16] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 655–687, 2008.

[17] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[18] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[19] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, pp. 253–263, Jan. 2008.

[20] A. Feuer and A. Nemirovski, "On sparse representations in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1579–1581, Jun. 2003.

[21] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, 1974.

[22] G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*. Oxford, UK: Oxford University Press, 2006.

[23] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel, "Z_4-Kerdock codes, orthogonal spreads, and extremal Euclidean line-sets," *Proc. London Math. Soc. (3)*, vol. 75, no. 2, pp. 436–480, 1997.

[24] T. Strohmer and R. W. Heath Jr., "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, vol. 14, pp. 257–275, 2003.

[25] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *Appl. Comput. Harmon. Anal.*, Feb. 2011, submitted.

[26] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," available online: http://www.acm.caltech.edu/jtropp/reports/Tro03-Greed-Good-TR.pdf, Feb. 2003, technical report.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge Press, 1985.